

DOCUMENT RESUME

ED 195 557

TM 800 775

AUTHOR Anderson, Scarvia B.
 TITLE The Self-Conscious Evaluator.
 PUB DATE Nov 80
 NOTE 13p.: Paper presented as a Presidential address, Evaluation Research Society (Washington, DC, November 20, 1980).
 EDRS PRICE MF01/PC01 Plus Postage.
 DESCRIPTORS *Evaluation Methods: *Evaluative Thinking: *Evaluators: *Experimenter Characteristics: *Program Evaluation

ABSTRACT

Some of the most important evaluation questions cannot be answered through experimental research. The "assessment" of the St. Louis Leadership Program (SLLP), sponsored by the Danforth Foundations, is used as an illustration of a "connoisseur-based study" approach to evaluation. The project team, given limited time and funds, decided that it could not determine the effectiveness of the SLLP per se. However, it could and did do the following tasks in assessing the quality of SLLP: (1) ask participants for their opinions about the program, interview Danforth staff, examine program files, and observe aspects of the 1979-80 program in operation; (2) describe the selection and other program processes from an external perspective; (3) analyze interrelationships among program components; (4) interpret the findings; and (5) make recommendations based on interpretation of the findings. Through observation, description, and judgment, some defensible and useful advice was given, but was it research? Judgment, rather than some of the methods and standard approaches, must be given a central place in evaluation. (RL)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

Anderson

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

The Self-Conscious Evaluator *

Scarvia B. Anderson
Educational Testing Service

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRO-
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN-
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT
OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY

Program evaluation is a relatively new field qua field. Oh, of course, people have been evaluating programs for a long time--in education, in the military, for health care, for harbor use. But I wrote the words 'program evaluation' on a piece of paper for the first time in 1965. This Society is only in its fourth year. Requirements for evaluation of federal programs are still an innovation. Few of us in this room started out to be program evaluators. We planned careers in the social sciences, operations research, and similar areas.

Those of us who came to evaluation from social science learned early on that some of the most important evaluation questions cannot be answered through experimental research. We also learned that conditions frequently dictated against decent experiments even when they were needed. We praised our new-found pragmatism: "Evaluators do what they have to do to answer the questions posed about a program, within technical and ethical boundaries." We declared our impatience "with the continuing argument over whether evaluation is or is not research" (Anderson and Ball, 1978, p. 9). But some of us--I daresay many of us--still felt a mite uncomfortable underneath. It was a "struggle," as Hansen has said, for evaluators "to preserve their heritage of experimental research despite the dictates of logic and lessons of experience" (1978, p. 15).

Let me share a recent personal experience with you. Last fall I was asked to chair a team to "assess" the St. Louis Leadership Program, sponsored by the Danforth Foundation. The situation included these elements:

- The choice of the other members of the team was theirs.
- The program was already into its fourth year of operation.

* Presidential address, Evaluation Research Society, Washington, D.C., November 20, 1980.

ED195557

TM 800 775

- An assessment of the effectiveness of the program during the fourth year had been recommended by the Board of Trustees.
- The decisions which the assessment was to inform were not specified.
- The report was due in about six months.
- The total allocation for the assessment was \$10,000.

I had just been working hard on the ERS Standards Committee, and the draft of the standards distributed to you this spring was well underway (1980). Among its 55 specifications are these:

- Evaluation staff should be selected, trained, and supervised to ensure competence and impartiality. (No. 21)
- For impact studies, the central evaluation design problem of estimating the effects of nontreatment should be fully described and justified. (No. 14)
- The information needs and expectations of clients for an evaluation should be made clear in advance. (No. 2)
- The feasibility of undertaking the evaluation effort should be estimated. (No. 6)
- The estimate of cost of a proposed effort should be prudent and based on sound accounting principles. (No. 4)
- Evaluators should not accept obligations that exceed their professional qualifications or the resources available to them. (No. 12)

The prospects for science or even reason in the St. Louis project were dim. I waved my pragmatism pennant and accepted the assignment. After all, I had been making the plea to other evaluators to do what has to be done to answer questions posed by important audiences. (Furthermore, we all have an obligation to the future employment of ERS members.) Seriously, we can preach the need for evaluation all we want to, and extoll the utility of the results

we are prepared to produce, but we will ultimately obtain our converts--and clients--by performing real services for them.

At our first meeting in St. Louis, the evaluation team got acquainted: a specialist in urban affairs and policy analysis, a director of leadership programs, an experimental psychologist turned measurement person and administrator, and my assistant whose field is business administration. We had done our homework, reading through stacks of materials about why the St. Louis Leadership Program was set up and what it hoped to accomplish.

The SLLP (as they call it) was initiated in 1976 as "an approach towards enhancing leadership development among selected leaders in metropolitan St. Louis in key policy-making positions and/or positions of influence....Participants give the equivalent of two full days each month to planned activities, as well as time for independent study." The objectives....are "to identify committed leaders from various...segments of the community, to expand the potential of community leaders for civic responsibilities and leadership..., to extend communication among these persons [and] with other community leaders..., and to prepare participants to work [effectively] within their own communities and in the metropolitan area..."

(Anderson, Drucker, Robinson, & Hill, 1980, pp. 1-3)

The first meeting in St. Louis provided the project team with the occasion to ask the Danforth staff what they thought the objectives and philosophy of the program really were and how it operated in practice.

We then turned to our major task, defining what we could do and what we couldn't do in assessing the quality of the St. Louis Leadership Program. We worded our concerns more elegantly in the final report than we were able to

at that first encounter:

At the most basic level, we could not "determine the effectiveness of the program" per se. That is, we could not evaluate it in the conventional sense: we did not have the time or funds to pursue adequate independent means of measuring the effects that the program was designed to accomplish; and, even if such measurement were possible, the constraints in the situation were such that we had no way of attributing the effects to the program....

What we could do was ask participants for their opinions about the program, interview Danforth staff, examine program files, and observe aspects of the 1979-80 program in operation. As a result, we could describe the selection and other program processes from an external perspective, analyze interrelationships among program components, interpret the findings and make recommendations in the context of the knowledge and experience we brought to the undertaking.

(Anderson, et al., 1980, p. x)

My God, we were connoisseur evaluators! Unfortunately, we did not have the comfort of knowing that at the time. The Stufflebeam-Webster (1980) classification of alternative approaches to evaluation had not been published. (A "connoisseur-based study" is one of the six categories of value-oriented studies they list.) But even though we didn't know "officially" what kind of approach we were taking, our misgivings subsided as we actually got down to work.

We divided that work up. I would write a general description of the program and survey past and present participants. Mark Drucker, the policy analyst,

would review the selection, nomination, and recruitment processes; study the operation of the program and interview a number of participants; and organize his own extensive knowledge of the climate for leadership in St. Louis. Frank Robinson, President of the Leadership Council of America, would also observe some program sessions, as well as compare the SLLP with other prominent leadership programs around the country. Robert Hill would analyze the demographic characteristics of accepted and rejected applicants over the four years of the program.

One of the problems that plagues all evaluators with limited time and resources--which means all evaluators--is the choice of things to look at. Among the many things we wanted to do in St. Louis, but couldn't, were to interview civic leaders and the employers of participants to find out what they thought about the SLLP, and to arrive at an operational definition of the term 'cultural diversity' which loomed so large but vaguely in discussions of the program. It is certainly not a question of right or wrong choices, but evaluators agonize over whether their choices are good, better, or best. Hersey, commenting on "dirty" research in "real" places says that "the decision about what to investigate represents a combination of intuition, political judgement, and value choice that is quite distinct from the objectivity of the research methods used to investigate a problem once it has been identified" (1979, p. 153). There is also a lot of luck and accident in the process. We did check from time to time with the Danforth folks to see if the things we were finding out about were things they wanted to know about.

Of course, the four of us did not work entirely independently. We got together from time to time to exchange ideas and observations and to lecture each other from our peculiar perspectives. They were good times and wonderful learning experiences for me--a side-benefit that evaluators should not view lightly.

Let me give you some examples of the kinds of information we reported to the Foundation:

- The SLLP, in line with one of its major objectives, is selecting a diverse group of participants from various ethnic, occupational, and other segments of the community. However, there are some segments of the community that appear to be underrepresented [and they are]....
- The major confusion in the selection process appears to be in whether criteria should be oriented toward "emergent" or [established] leaders.
- Participants from [the four] different years of the program are in substantial agreement about what the major objectives should be: create a network of future area leaders, increase knowledge about St. Louis, produce more involvement in community affairs, and improve leadership skills.
- Participants rate the program as generally successful in meeting the first three of [those] objectives, less successful in improving leadership skills.
- Many participants report positive changes in their jobs, income level, community involvement, and personal characteristics since they entered the program, although these changes did not necessarily result from participation in the SLLP.
- The values that current and past participants hold about leadership, decision-making, minority participation, and related topics are highly variable within and among groups.
- Considerable self-initiated communication, the basis for "networking," seems to be occurring among past participants.... (Anderson, et al., 1980, pp. iv, v)

We provided the Danforth directors and staff with an analysis of competing leadership strategies against which they could examine their program, and we offered a series of recommendations that followed from the information we obtained about the selection process, the design and implementation of the program, follow-up activities, and program management and support. The report ended with a predictable argument to the Foundation to make a suitable investment in the development of outcome measures for future training and monitoring purposes.

We were reasonably proud of ourselves. Through observation, description, and judgement we had given our client some defensible advice and they seemed to find it useful. But will those niggles, "Is it research?" "Is it professional?", return the next time an SLLP comes along? Will we be just a little embarrassed or apologetic about showing our reports to our colleagues from similar social science backgrounds?

Even the patently unselfconscious Glass takes time to protest that he is not "against high-quality research. After all," he says, "I was trained in the very crucible of research criticism, at Wisconsin, where we were taught first, that there are dozens of things one must do properly to achieve a valid study, and second, that they were impossible to do" (1979, p. 12).

Light holds out for experimental research although, in the context of synthesizing data across studies, he recognizes that "in a real world setting, one may be restricted to sets of data where 'nature' has dealt out the cards, and we do the best we can interpreting the hand" (1979, p. 8). Freeman and Solomon predict that in the next decade, "experiments and quasi-experiments will continue to be the methods of choice for the study of innovations, initiatives, and procedural and administrative changes. But the press, consistent with the sunset law movement and the new 'Proposition 13' mentality, is to use evaluations to assess the impact of ongoing programs" (1979, p. 259).

Freeman and Solomon seem to contrast experimentation and evaluation and certainly few evaluators actually do experiments. They subscribe, at least publicly, to the argument that "the political and social context of evaluation research is such that it is virtually impossible to operationalize the features of programs or to reach some agreement on program objectives, much less to measure those objectives in any methodologically respectable way. In short, the kind of control necessary for doing classical experiments is viewed as unattainable." I am quoting from Tornatzky (1979, p. 111) and he does not buy this argument, but he goes on to say that evaluators must be good at field experimentation and be influential scientist/administrators, advocates for exemplary programs, and involved in affecting social change (p. 115). Wiener (1979) takes a step further by suggesting that:

The more a professional evaluator can adapt to the managerial style, the more likely will evaluation research be utilized. However, the more the professional evaluator adapts to the managerial style, the less he or she will be able to adhere, for example, to the professional norm of competency (which at a minimum requires much more than casual regard for the issue of validity) and to the norms of altruism and service. (p. 233)

Wiener finds resolution of the evaluation-as-profession issue not in evaluation becoming a profession but in evaluation becoming a specialty of the professional manager!

Once in the midst of the St. Louis project I remembered an apartment my mother and I had with an old bathtub that sat up on legs in the middle of the room. All my classmates' houses seemed to have colored ceramic tubs neatly set between tiled walls. So what did Mother do with our tub? Painted roses all over it and gilded its claw feet. "If you can't hide it, make a feature of it," she always said.

Some evaluators have done that, sometimes very effectively. Others seem to forget which measures and approaches were conceived in this philosophy and tend to perpetuate them out of context and out of proportion to their worth. We can attribute at least some of the proliferation of evaluation types in the literature to these evaluators. But there are other motives leading to the long list that includes responsive, transactional, discrepancy, phenomenological, goal-free, meta, developmental, utilization, impact--far too many to name here. The closet experimentalists are probably responsible for some of them. There is something eminently respectable about naming what you are doing. How much nicer to say "I am doing a naturalistic evaluation" than "I'm just mucking about." And I've already told you about the great feeling of satisfaction derived from labeling oneself "a connoisseur evaluator." In all fairness, some of the labels have been invented because evaluation is something different and its concepts are not captured by borrowed terms. And the explications have helped greatly to clarify the diverse nature of our field.

But the question I have about many of these evaluation approaches is "Once we've named it, do we have to feed it? Does it own us?" Sam Ball and I once suggested that researchers fall into two categories: the method-bound and the problem-oriented.

The method-bound have a favorite method, analytic technique, or instrument and spend their time looking for a problem to apply it to; if their search fails, they may apply the method anyway--either inappropriately or gratuitously. The problem-oriented...may also occasionally apply methodologies inappropriately--but at least they start with the premise that the choice of methodology should follow from, not precede, the delineation of the purpose of the investigation.

(1978, p. 43)

Can--will--Stake do a "nonresponsive" evaluation? How about a survey testing project if the occasion calls for it? Can Kiresuk and his colleagues attack an evaluation problem that is not amenable to Goal Attainment Scaling? Would Stone undertake--or encourage his students to undertake--a comparative experimental study of competing treatments? Will method- or approach-independence signal evaluation's coming of age?

That would seem to be one characteristic of Windle's definition of the third or most mature stage in the evolution of program evaluation. He describes this stage in terms of accountability: "Accountability in program evaluation means enabling others to share control over the program by making trustworthy information about the program available" (1979, p. 195).

According to Windle, in the first stage:

New evaluators assume uncritically and idealistically the long term rationale of basic research, that production of knowledge will lead to improvements in practice....The second stage is the reaction of cynicism which results from recognizing and being frustrated by the politics in evaluation. (pp. 194-195)

I think my remarks today indicate that I am still in this second stage--frustrated as much by the professional politics I fear evaluators play as by the politics inherent in the evaluation of social systems. But I also think that the kind of self-examination I am engaging in today--and that I am urging on you--can move us forward. In a few years, I hope that such introspection will be excessive, that we will have enough confidence in what we do that we don't worry so much about what we are.

To develop that confidence, we will not only have to gain independence from some of our methods and standard approaches and give a central place to judgment, but also cast off the feelings of inadequacy that have so often led to pretentious invention, recognize that by and large we are dealing with unpredictable systems and organize our responses accordingly, and not take

ourselves too seriously. Then in Glass's (1979, p. 14) words, we may begin to command, and help our clients command, options rather than flounder around in an ineffectual search for eternal truth. Or, as in Noel Coward's creed, we can believe in doing what we can, crying when we must, and laughing when we choose.

References

- Anderson, S. B., & Ball, S. The profession and practice of program evaluation. San Francisco: Jossey-Bass, 1978.
- Anderson, S. B., Drucker, M. L., Robinson, F., & Hill, R. B. St. Louis Leadership Program--description, analysis, and recommendations. Atlanta: Educational Testing Service, 1980.
- Evaluation Research Society. Standards for program evaluation (Exposure draft). Washington, D. C.: Author, 1980.
- Freeman, H. E., & Solomon, M. A. The next decade in evaluation research. Evaluation and Program Planning, 1979, 2, 255-262.
- Glass, G. V. Policy for the unpredictable (Uncertainty research and policy). Educational Researcher, 1979, 8(9), 12-14.
- Hansen, J. B. The formative-summative dichotomy and the role of description in program evaluation. Evaluation News, no. 6, August 1978, 14-16.
- Hersey, J. C. "Dirty" research in "real" places: A practitioner's guide to program evaluation in the human services. Evaluation and Program Planning, 1979, 2, 153-157.
- Light, R. J. Capitalizing on variation: How conflicting research findings can be helpful for policy. Educational Researcher, 1979, 8(9), 7-11.
- Stufflebeam, D. L., & Webster, W. J. An analysis of alternative approaches to evaluation. Educational Evaluation and Policy Analysis, 1980, 2(3), 5-20.
- Tornatsky, L. G. The triple-threat evaluator. Evaluation and Program Planning, 1979, 2, 111-115.
- Wiener, A. The development of evaluation as a concession. Evaluation and Program Planning, 1979, 2, 231-234.
- Windle, C. Developmental trends in program evaluation. Evaluation and Program Planning, 1979, 2, 193-196.